A robust all-at-once multigrid method for the Stokes control problem

Stefan Takacs

Abstract In this paper we present an all-at-once multigrid method for a distributed Stokes control problem (velocity tracking problem). For solving such a problem, we use the fact that the solution is characterized by the optimality system (Karush-Kuhn-Tucker-system). The discretized optimality system is a large-scale linear system whose condition number depends on the grid size and on the choice of the regularization parameter forming a part of the problem. Recently, block-diagonal preconditioners have been proposed, which allow to solve the problem using a Krylov space method with convergence rates that are robust in both, the grid size and the regularization parameter or cost parameter. In the present paper, we develop an all-at-once multigrid method for a Stokes control problem and show robust convergence, more precisely, we show that the method converges with rates which are bounded away from one by a constant which is independent of the grid size and the choice of the regularization or cost parameter.

1 Introduction

In the present paper, we consider the following model problem *(distributed velocity tracking problem)*. Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with $d \in \{2,3\}$. Find a velocity field $u \in [H^1(\Omega)]^d$, a pressure distribution $p \in L^2(\Omega)$ and a

Stefan Takacs
Visiting Postdoc, Mathematical Institute, University of Oxford, United Kingdom
Tel.: +44-1865-(6)15312
Fax: +44-1865-273583
E-mail: stefan.takacs@maths.ox.ac.uk
E-mail: stefan.takacs@numa.uni-linz.ac.at

control (force field) $f \in [L^2(\Omega)]^d$ such that the tracking functional

$$J(u, f) = \frac{1}{2} \|u - u_D\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|f\|_{L^2(\Omega)}^2$$

is minimized subject to the Stokes equations

$$
\begin{aligned}
-\Delta u + \nabla p &= f \text{ in } \Omega, \\
\nabla \cdot u &= 0 \text{ in } \Omega, \\
u &= 0 \text{ on } \partial\Omega.
\end{aligned}
$$

The cost parameter or regularization parameter $\alpha > 0$ and the desired state (desired velocity field) $u_D \in [L^2(\Omega)]^d$ are assumed to be given. To enforce uniqueness of the solution, we additionally require $\int_\Omega p \, \mathrm{d}x = 0$.

Here and in what follows, $L^2(\Omega)$ and $H^1(\Omega)$ denote the standard Lebesgue and Sobolev spaces with associated standard norms $\| \cdot \|_{L^2(\Omega)}$ and $\| \cdot \|_{H^1(\Omega)}$, respectively.

The main goal of this work is to construct and to analyze numerical methods that produce an approximate solution to the optimization problem, where the computational complexity can be bounded by the number of unknowns times a constant which is independent of the grid level and the choice of the parameter α, in particular for small values of α.

The solution of the optimization problem is characterized by the Karush-Kuhn-Tucker-system (KKT-system). As we are interested in good approximations of the solution, the discretization of the KKT-system leads to a large-scale linear system. This linear system will be solved with multigrid methods because they are one of the fastest known methods for such problems. Originally, multigrid methods have been designed and analyzed for elliptic problems. They also work well for saddle point problems (like KKT-systems) and have gained growing interest in this area, see, e.g., [4] and the references cited there.

The unknowns of the discretized KKT-system for a PDE-constrained optimization problem can be partitioned into primal variables and dual variables. In our case, the primal variables consist of state variables (the velocity field u and the pressure distribution p) and control variables (the force field f). The dual variables are the Lagrange multipliers which are introduced to incorporate the constraints. One approach to solve such problems is to apply multigrid methods in every step of an overall block-structured iterative method to equations in just one of these blocks of variables. Such methods have been proposed, e.g., in [18,9,12].

Another approach, which we will follow here, is to apply the multigrid idea directly to the (reduced or not reduced) KKT-system, which is called an all-at-once approach. Such methods have been proposed and discussed for the elliptic optimal control problem, e.g., in [13,16].

In this paper we present a convergence proof for multigrid methods based on the classical splitting of the analysis into smoothing property and approximation property, see [7].

This paper is organized as follows. In Section 2 we will introduce the optimality system and discuss its discretization. In Section 3 we will introduce an all-at-once multigrid approach. Its convergence will be proven in Section 4. Numerical results which illustrate the convergence result will be presented in Section 5. In Section 6 we will close with conclusions.

2 Optimality system and discretization

For setting up the optimality system we need the space $H_0^1(\Omega)$, the space of functions in $H^1(\Omega)$ vanishing on the boundary. Moreover, we need the space $L_0^2(\Omega)$, which is the space of functions in $L^2(\Omega)$ with mean value 0, i.e.,

$$L_0^2(\Omega) := \left\{ v \in L^2(\Omega) \ : \ \int_\Omega v \, \mathrm{d}\xi = 0 \right\}.$$

Both spaces are equipped with standard norms, i.e., $\| \cdot \|_{H_0^1(\Omega)} := \| \cdot \|_{H^1(\Omega)}$ and $\| \cdot \|_{L_0^2(\Omega)} := \| \cdot \|_{L^2(\Omega)}$.

The solution of the problem is characterized by the Karush Kuhn Tucker system (KKT-system), which reads as follows, cf. [18] and others.

Find $(u, p, f, \lambda, \mu) \in [H_0^1(\Omega)]^d \times L_0^2(\Omega) \times [L^2(\Omega)]^d \times [H_0^1(\Omega)]^d \times L_0^2(\Omega)$ such that

$$(u, \tilde{u})_{L^2(\Omega)} + (\nabla \lambda, \nabla \tilde{u})_{L^2(\Omega)} + (\mu, \nabla \cdot \tilde{u})_{L^2(\Omega)} = (u_D, \tilde{u})_{L^2(\Omega)}$$
$$(\nabla \cdot \lambda, \tilde{p})_{L^2(\Omega)} = 0$$
$$\alpha(f, \tilde{f})_{L^2(\Omega)} - (\lambda, \tilde{f})_{L^2(\Omega)} = 0$$
$$(\nabla u, \nabla \tilde{\lambda})_{L^2(\Omega)} + (p, \nabla \cdot \tilde{\lambda})_{L^2(\Omega)} - (f, \tilde{\lambda})_{L^2(\Omega)} = 0$$
$$(\nabla \cdot u, \tilde{\mu})_{L^2(\Omega)} = 0$$

holds for all $(\tilde{u}, \tilde{p}, \tilde{f}, \tilde{\lambda}, \tilde{\mu}) \in [H_0^1(\Omega)]^d \times L_0^2(\Omega) \times [L^2(\Omega)]^d \times [H_0^1(\Omega)]^d \times L_0^2(\Omega)$.

The third line of the KKT-system directly implies $f = \alpha^{-1}\lambda$.

This allows to eliminate the control f, which leads to the reduced KKT-system, which reads as follows. Find $x := (u, p, \lambda, \mu) \in X := [H_0^1(\Omega)]^d \times L_0^2(\Omega) \times [H_0^1(\Omega)]^d \times L_0^2(\Omega)$ such that

$$(u, \tilde{u})_{L^2(\Omega)} + (\nabla \lambda, \nabla \tilde{u})_{L^2(\Omega)} + (\mu, \nabla \cdot \tilde{u})_{L^2(\Omega)} = (u_D, \tilde{u})_{L^2(\Omega)}$$
$$(\nabla \cdot \lambda, \tilde{p})_{L^2(\Omega)} = 0$$
$$(\nabla u, \nabla \tilde{\lambda})_{L^2(\Omega)} + (p, \nabla \cdot \tilde{\lambda})_{L^2(\Omega)} - \alpha^{-1}(\lambda, \tilde{\lambda})_{L^2(\Omega)} = 0$$
$$(\nabla \cdot u, \tilde{\mu})_{L^2(\Omega)} = 0$$

holds for all $(\tilde{u}, \tilde{p}, \tilde{\lambda}, \tilde{\mu}) \in X$.

Certainly, the KKT-system can be rewritten as one variational equation as follows. Find $x \in X$ such that

$$\mathcal{B}(x, \tilde{x}) = \mathcal{F}(\tilde{x}) \qquad \text{for all } \tilde{x} \in X, \tag{1}$$

where

$$\mathcal{B}((u, p, \lambda, \mu), (\tilde{u}, \tilde{p}, \tilde{\lambda}, \tilde{\mu})) := (u, \tilde{u})_{L^2(\Omega)} + (\nabla\lambda, \nabla\tilde{u})_{L^2(\Omega)} + (\mu, \nabla \cdot \tilde{u})_{L^2(\Omega)}$$
$$+ (\nabla \cdot \lambda, \tilde{p})_{L^2(\Omega)} + (\nabla u, \nabla\tilde{\lambda})_{L^2(\Omega)} + (p, \nabla \cdot \tilde{\lambda})_{L^2(\Omega)} - \alpha^{-1}(\lambda, \tilde{\lambda})_{L^2(\Omega)}$$
$$+ (\nabla \cdot u, \tilde{\mu})_{L^2(\Omega)} \text{ and}$$
$$\mathcal{F}(\tilde{u}, \tilde{p}, \tilde{\lambda}, \tilde{\mu}) := (u_D, \tilde{u})_{L^2(\Omega)}.$$

We are interested in finding an approximative solution for equation (1). Both, the proposed solution strategy and the convergence analysis, follow the abstract framework introduced in [16]. The conditions, **(A1)**, **(A1a)**, **(A3)** and **(A4)**, mentioned in the present paper are the same conditions as in [16].

For simplicity, we introduce the following notation.

Notation 1 *Throughout this paper, $C > 0$ is a generic constant, independent of the grid level k and the choice of the parameter α. For any scalars a and b, we write $a \lesssim b$ (or $b \gtrsim a$) if there is a constant $C > 0$ such that $a < C\,b$. We write $a \eqsim b$ if $a \lesssim b \lesssim a$.*

The following property guarantees existence and uniqueness of the solution.

(A1) The relation

$$\|x\|_X \lesssim \sup_{0 \neq \tilde{x} \in X} \frac{\mathcal{B}(x, \tilde{x})}{\|\tilde{x}\|_X} \lesssim \|x\|_X$$

holds for all $x \in X$.

In [18] it was shown that condition **(A1)** is satisfied for $X := Y \times Y$, where $Y := U \times P$, $U := [H_0^1(\Omega)]^d$, $P := L_0^2(\Omega)$, equipped with norms

$$\|x\|_X^2 := \|(u, p, \lambda, \mu)\|_X^2 := \|(u, p)\|_Y^2 + \alpha^{-1}\|(\lambda, \mu)\|_Y^2,$$

where

$$\|(u, p)\|_Y^2 := \|u\|_U^2 + \|p\|_P^2,$$
$$\|u\|_U^2 := \|u\|_{L^2(\Omega)}^2 + \alpha^{1/2}\|u\|_{H^1(\Omega)}^2 \text{ and}$$
$$\|p\|_P^2 := \sup_{0 \neq w \in [H_0^1(\Omega)]^d} \frac{\alpha(p, \nabla \cdot w)_{L^2(\Omega)}^2}{\|w\|_{L^2(\Omega)}^2 + \alpha^{1/2}\|w\|_{H^1(\Omega)}^2}.$$

Using the following notation, we can express the norms in a nicer way.

Notation 2 *For any Hilbert space A, the symbol A^* denotes its dual space equipped with the dual norm*

$$\|u\|_{A^*} := \sup_{0 \neq w \in A} \frac{\langle u, w \rangle}{\|w\|_A},$$

where $\langle u, \cdot \rangle := u(w)$ denotes the duality pairing.

For any Hilbert space A and any scalar $a > 0$, the symbol $a\,A$ denotes the space on the underlying set of the Hilbert space A equipped with the norm

$$\|u\|_{a\,A}^2 := a\|u\|_A^2.$$

For any two Hilbert spaces A and B, the symbol $A \cap B$ denotes the space on the intersection of the underlying sets, $\{u \in A \cap B\}$, equipped with the norm

$$\|u\|_{A \cap B}^2 := \|u\|_A^2 + \|u\|_B^2,$$

and the symbol $A + B$ denotes the space on the algebraic sum of the underlying sets, $\{u_1 + u_2 \ : \ u_1 \in A, u_2 \in B\}$, equipped with the norm

$$\|u\|_{A+B}^2 := \inf_{u_1 \in A, u_2 \in B, u = u_1 + u_2} \|u_1\|_A^2 + \|u_2\|_B^2.$$

The spaces A^*, $a\,A$, $A \cap B$ and $A + B$ are Hilbert spaces. The fact that A^* is a Hilbert space follows directly from the Riesz representation theorem, see, e.g., Theorem 1.2 in [1]. The fact that $a\,A$ is a Hilbert space is obvious and for the latter two see, e.g., Lemma 2.3.1 in [3].

We immediately see, that the norm on U can be rewritten as follows

$$\|u\|_U = \|u\|_{L^2(\Omega) \cap \alpha^{1/2} H^1(\Omega)}.$$

To reformulate the norm $\|\cdot\|_P$, we need a regularity assumption.

(R) *Regularity of the generalized Stokes problem.* Let $f \in [L^2(\Omega)]^d$ and $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$ be arbitrarily but fixed and $(u, p) \in [H_0^1(\Omega)]^d \times L_0^2(\Omega)$ be the solution of the Stokes problem, i.e., such that

$$\begin{aligned}
(\nabla u, \nabla \tilde{u})_{L^2(\Omega)} + (p, \nabla \cdot \tilde{u})_{L^2(\Omega)} &= (f, \tilde{u})_{L^2(\Omega)} \\
(\nabla \cdot u, \tilde{p})_{L^2(\Omega)} &= (g, \tilde{p})_{L^2(\Omega)}
\end{aligned}$$

holds for all $(\tilde{u}, \tilde{p}) \in [H_0^1(\Omega)]^d \times L_0^2(\Omega)$.
Then $(u, p) \in [H^2(\Omega)]^d \times H^1(\Omega)$ and

$$\|u\|_{H^2(\Omega)}^2 + \|p\|_{H^1(\Omega)}^2 \lesssim \|f\|_{L^2(\Omega)}^2 + \|g\|_{H^1(\Omega)}^2.$$

This condition is satisfied for convex polygonal domains, see Lemma 2.1 in [14] which is a direct consequence of Theorem 2 in [8].

Lemma 1 *If **(R)** is satisfied, then*

$$\|p\|_P \eqsim \|p\|_{\alpha H^1(\Omega) + \alpha^{1/2} L^2(\Omega)}$$

holds for all $p \in L_0^2(\Omega)$.

This lemma was shown in Theorem 3.2 in [11] under a regularity assumption, which is weaker than regularity assumption **(R)**.

The discretization of problem (1) is done using standard finite element techniques. We assume to have a sequence of girds obtained by uniform refinement. On each grid level k, we discretize the problem using the Galerkin approach, i.e., we have finite dimensional spaces $X_k \subseteq X$ and consider the following problem. Find $x_k \in X_k$ such that

$$\mathcal{B}(x_k, \tilde{x}_k) = \mathcal{F}(\tilde{x}_k) \qquad \text{for all } \tilde{x}_k \in X_k. \tag{2}$$

Using a nodal basis, we can represent this problem in matrix-vector notation as follows:

$$\mathcal{A}_k \underline{x}_k = \underline{f}_k. \tag{3}$$

Here and in what follows, any underlined quantity, like $\underline{x}_k$, is the representation of the corresponding non-underlined quantity, here x_k, with respect to a nodal basis of the corresponding Hilbert space, here X_k.

Existence and uniqueness of the discretized problem is guaranteed by the following condition.

(A1a) The relation

$$\|x_k\|_X \lesssim \sup_{0 \neq \tilde{x}_k \in X_k} \frac{\mathcal{B}(x_k, \tilde{x}_k)}{\|\tilde{x}_k\|_X} \lesssim \|x_k\|_X$$

holds for all $x_k \in X_k$.

Due to the fact that the model problem is indefinite, condition **(A1)** does not imply condition **(A1a)**. For the Stokes problem itself, it is well-known that such a condition (also known as *discrete inf-sup condition*) can only be guaranteed if the discretization is chosen appropriately. The same is true for the Stokes control problem. Fortunately, we can show the discrete inf-sup condition **(A1a)** for the Stokes control problem based on pre-existing knowledge on the discrete inf-sup condition for the Stokes problem. This allows to show that all discretizations which are suitable for the Stokes flow problem are also suitable for the Stokes control problem.

We choose the space X_k as follows:

$$X_k := Y_k \times Y_k \qquad \text{where} \qquad Y_k := U_k \times P_k$$

and the choice of $U_k \subseteq U = [H_0^1(\Omega)]^d$ and $P_k \subseteq P = L_0^2(\Omega)$ is discussed below. Note that X_k has product structure and that the state and the adjoined state (Lagrange multipliers) are discretized the same way. The same has already been done for optimal control problems with elliptic state equation, cf. [16] and many others.

Due to the fact that the grids are obtained by uniform refinement, the discrete subsets are nested, i.e., $U_k \subseteq U_{k+1}$ and $P_k \subseteq P_{k+1}$. Therefore, also $X_k \subseteq X_{k+1}$ holds.

The next step is to show condition **(A1a)**. We have seen that the analysis done in [18], applied to the infinite dimensional spaces, shows condition **(A1)**. If the analysis done in [18] is applied to the discretized spaces, we obtain that

$$\|x_k\|_{X_k} \lesssim \sup_{0 \neq \tilde{x}_k \in X_k} \frac{\mathcal{B}(x_k, \tilde{x}_k)}{\|\tilde{x}_k\|_{X_k}} \lesssim \|x_k\|_{X_k} \tag{4}$$

is satisfied for all $x_k \in X_k$, where

$$\|x_k\|_{X_k}^2 := \|(u_k, p_k, \lambda_k, \mu_k)\|_{X_k}^2 := \|(u_k, p_k)\|_{Y_k}^2 + \alpha^{-1} \|(\lambda_k, \mu_k)\|_{Y_k}^2,$$

$$\|(u_k, p_k)\|_{Y_k}^2 := \|u_k\|_U^2 + \|p_k\|_{P_k}^2,$$

$$\|p_k\|_{P_k}^2 := \sup_{0 \neq w_k \in U_k} \frac{\alpha(p_k, \nabla \cdot w_k)_{L^2(\Omega)}^2}{\|w_k\|_{L^2(\Omega)}^2 + \alpha^{1/2} \|w_k\|_{H^1(\Omega)}^2} \quad \text{and}$$

$\|\cdot\|_U^2$ is as above.

Note that this is not condition **(A1a)**, as the norms $\|\cdot\|_P$ and $\|\cdot\|_{P_k}$ are not equal. For showing condition **(A1a)**, it suffices to show that these two norms are equivalent which implies also the equivalence of the norms $\|\cdot\|_X$ and $\|\cdot\|_{X_k}$. This can be shown using the following condition.

(S) The discretization of P is H^1-conforming, i.e., $P_k \subseteq H^1(\Omega)$, and the *weak inf-sup condition*

$$\sup_{0 \neq u_k \in U_k} \frac{(\nabla \cdot u_k, p_k)_{L^2(\Omega)}}{\|u_k\|_{L^2(\Omega)}} \gtrsim \|\nabla p_k\|_{L^2(\Omega)}$$

holds for all $p_k \in P_k$.

Lemma 2 *Assume that the discretization satisfies condition* **(S)**. *Then condition* **(A1a)** *is satisfied for the model problem.*

Proof Lemma 2.2 in [10] states (provided that **(S)** is satisfied) that $\|\cdot\|_P \eqsim \|\cdot\|_{P_k}$ is satisfied. A direct consequence is $\|\cdot\|_X \eqsim \|\cdot\|_{X_k}$. Therefore, condition (4) implies condition **(A1a)**. $\qquad\square$

Note that condition **(S)** is a standard condition which ensures that the chosen discretization is stable for the Stokes problem. In [2,17] it was shown that condition **(S)** is satisfied for the Taylor-Hood element ($P1 - P2$-element) for polygonal domains where at least one vertex of each element is located in the interior of the domain. Here and in what follows we assume that the problem is discretized with the Taylor-Hood element and that the mesh satisfies the named condition.

3 An all-at-once multigrid method

The problem shall be solved with an all-at-once multigrid method. The abstract algorithm for solving the discretized equation (3) on grid level k reads as follows. Starting from an initial approximation $\underline{x}_k^{(0)}$, one iterate of the multigrid method is given by the following two steps:

– *Smoothing procedure:* Compute

$$\underline{x}_k^{(0,m)} := \underline{x}_k^{(0,m-1)} + \hat{\mathcal{A}}_k^{-1} \left(\underline{f}_k - \mathcal{A}_k\, \underline{x}_k^{(0,m-1)} \right) \qquad \text{for } m = 1, \ldots, \nu$$

with $\underline{x}_k^{(0,0)} = \underline{x}_k^{(0)}$. The choice of the smoother (or, in other words, of the preconditioning matrix $\hat{\mathcal{A}}_k^{-1}$) will be discussed below.

– *Coarse-grid correction:*
  – Compute the defect $\underline{r}_k^{(1)} := \underline{f}_k - \mathcal{A}_k\, \underline{x}_k^{(0,\nu)}$ and restrict it to grid level $k - 1$ using a restriction matrix I_k^{k-1}:

  $$\underline{r}_{k-1}^{(1)} := I_k^{k-1} \left(\underline{f}_k - \mathcal{A}_k\, \underline{x}_k^{(0,\nu)} \right).$$

– Solve the coarse-grid problem

$$\mathcal{A}_{k-1}\,\underline{p}_{k-1}^{(1)} = \underline{r}_{k-1}^{(1)} \tag{5}$$

approximatively.
– Prolongate $\underline{p}_{k-1}$ to the grid level k using an prolongation matrix I_{k-1}^{k} and add the result to the previous iterate:

$$\underline{x}_{k}^{(1)} := \underline{x}_{k}^{(0,\nu)} + I_{k-1}^{k}\,\underline{p}_{k-1}^{(1)}.$$

As we have assumed to have nested spaces, the intergrid-transfer matrices I_{k-1}^{k} and I_{k}^{k-1} can be chosen in a canonical way: I_{k-1}^{k} is the canonical embedding and the restriction I_{k}^{k-1} is its transpose.

If the problem on the coarser grid is solved exactly (two-grid method), the coarse-grid correction is given by

$$\underline{x}_{k}^{(1)} := \underline{x}_{k}^{(0,\nu)} + I_{k-1}^{k}\,\mathcal{A}_{k-1}^{-1}\,I_{k}^{k-1}\left(\underline{f}_{k} - \mathcal{A}_{k}\,\underline{x}_{k}^{(0,\nu)}\right). \tag{6}$$

In practice the problem (5) is approximatively solved by applying one step (V-cycle) or two steps (W-cycle) of the multigrid method, recursively. On the coarsest grid level ($k = 0$) the problem (5) is solved exactly.

To construct a multigrid convergence result based on Hackbusch's splitting of the analysis into smoothing property and approximation property, we have to introduce an appropriate framework.

Convergence is shown on the spaces X_k, which are equipped with an L^2-like norms $\|\!|\cdot|\!\|_{0,k}$, which are defined a follows:

$$\|\!|x_k|\!\|_{0,k}^2 := \|\underline{x}_k\|_{\mathcal{L}_k}^2 := (\mathcal{L}_k\underline{x}_k,\underline{x}_k)_{\ell^2},$$

where

$$\mathcal{L}_k := \begin{pmatrix} \varphi_{\alpha,k}M_{U,k} & & & \\ & \alpha h_k^{-2}\varphi_{\alpha,k}^{-1}M_{P,k} & & \\ & & \alpha^{-1}\varphi_{\alpha,k}M_{U,k} & \\ & & & h_k^{-2}\varphi_{\alpha,k}^{-1}M_{P,k} \end{pmatrix}, \tag{7}$$

and $\varphi_{\alpha,k} := 1 + \alpha^{1/2}h_k^{-2}$ and $M_{U,k}$ and $M_{P,k}$ are the mass matrices, representing the L^2-inner product on U_k and P_k, respectively. Based on the norm $\|\!|\cdot|\!\|_{0,k}$, we can introduce the residual norm $\|\!|\cdot|\!\|_{2,k}$ using

$$\|\!|x_k|\!\|_{2,k} := \sup_{\tilde{x}_k \in X_k} \frac{\mathcal{B}\left(x_k, \tilde{x}_k\right)}{\|\!|\tilde{x}_k|\!\|_{0,k}}.$$

Smoothing property and approximation property read as follows.

– *Smoothing property:*

$$\|\!|x_k^{(0,\nu)} - x_k^*|\!\|_{2,k} \le \eta(\nu)\|\!|x_k^{(0)} - x_k^*|\!\|_{0,k} \tag{8}$$

should hold for some function $\eta(\nu)$ with $\lim_{\nu\to\infty}\eta(\nu) = 0$. Here and in what follows, $x_k^* \in X_k$ is the exact solution of the discretized problem (3).

– *Approximation property:*

$$\||x_k^{(1)} - x_k^*\||_{0,k} \leq C_A \||x_k^{(0,\nu)} - x_k^*\||_{2,k} \tag{9}$$

should hold for some constant $C_A > 0$.

It is easy to see that, if we combine both conditions, we obtain

$$\||x_k^{(1)} - x_k^*\||_{0,k} \leq q(\nu)\||x_k^{(0)} - x_k^*\||_{0,k},$$

where $q(\nu) = C_A \eta(\nu)$, i.e., that the two-grid method converges for ν large enough. The convergence of the W-cycle multigrid method can be shown under mild assumptions, see e.g. [7].

The choice of an appropriate smoother is a key issue in constructing such a multigrid method. Here, we introduce one smoother which is appropriate for a large class of problems including the model problem: *normal equation smoothers*, cf. [5], which read as follows.

$$\underline{x}_k^{(0,m)} := \underline{x}_k^{(0,m-1)} + \tau \underbrace{\mathcal{L}_k^{-1} \mathcal{A}_k \mathcal{L}_k^{-1}}_{\hat{\mathcal{A}}_k^{-1} :=} \left(\underline{f}_k - \mathcal{A}_k \, \underline{x}_k^{(0,m-1)} \right) \qquad \text{for } m = 1, \ldots, \nu.$$

Here, a fixed $\tau > 0$ has to be chosen such that the spectral radius $\rho(\tau \hat{\mathcal{A}}_k^{-1} \mathcal{A}_k)$ is bounded away from 2 on all grid levels k and for all choices of the parameter α.

Using a standard inverse inequality, one can show that

$$\|x_k\|_X \lesssim \||x_k\||_{0,k}$$

is satisfied for all $x_k \in X_k$. Based on this result, using an eigenvalue analysis one can show the following lemma, cf. [5].

Lemma 3 *The damping parameter $\tau > 0$ can be chosen independent of grid level k and the choice of the parameter α such that*

$$\tau \rho(\hat{\mathcal{A}}_k^{-1} \mathcal{A}_k) \leq 2 - \epsilon < 2,$$

holds for some constant $\epsilon > 0$. For this choice of τ, there is a constant $C_S > 0$, independent of the grid level k and the choice of the parameter α, such that the smoothing property (8) *is satisfied with rate*

$$\eta(\nu) := C_S \nu^{-1/2}.$$

Certainly, the iteration procedure (8) should be efficient-to-apply. Using the fact, that the mass matrices $M_{U,k}$ and $M_{P,k}$ in (7) and their diagonals are spectrally equivalent under weak assumptions, for the practical realization of the smoother these matrices can be replaced by their diagonals.

4 A convergence proof

The proof of the approximation property is done using the approximation theorem introduced in [16] which requires besides the conditions **(A1)** and **(A1a)** two more conditions (conditions **(A3)** and **(A4)**) involving, besides the Hilbert space X, two more Hilbert spaces $X_{-,k} := (X_-, \| \cdot \|_{X_{-,k}})$ and $X_{+,k} := (X_+, \| \cdot \|_{X_{+,k}})$, which are chosen as follows.

As weaker space, we choose $X_- := Y_- \times Y_-$, where $Y_- := U_- \times P_-$, $U_- := [L^2(\Omega)]^d$ and $P_- := [H_0^1(\Omega) \cap L_0^2(\Omega)]^*$, equipped with norms

$$\|x\|_{X_{-,k}}^2 := \|(u,p,\lambda,\mu)\|_{X_{-,k}}^2 := \|(u,p)\|_{Y_{-,k}}^2 + \alpha^{-1}\|(\lambda,\mu)\|_{Y_{-,k}}^2,$$

$$\|(u,p)\|_{Y_{-,k}}^2 := \|u\|_{U_{-,k}}^2 + \|p\|_{P_{-,k}}^2,$$

$$\|u\|_{U_{-,k}}^2 := h_k^{-2}\|u\|_{[H_0^1(\Omega)+\alpha^{-1/2}L^2(\Omega)]^*}^2 \text{ and}$$

$$\|p\|_{P_{-,k}}^2 := h_k^{-2}\|p\|_{[\alpha^{-1}L_0^2(\Omega)\cap\alpha^{-1/2}H_0^1(\Omega)]^*}^2$$

Note that dual spaces are $(X_-)^* := (Y_-)^* \times (Y_-)^*$, where $(Y_-)^* = (U_-)^* \times (P_-)^*$, $(U_-)^* = [L^2(\Omega)]^d$ and $(P_-)^* = H_0^1(\Omega) \cap L_0^2(\Omega)$, equipped with norms

$$\|\mathcal{F}\|_{(X_{-,k})^*}^2 := \|(f,g,\zeta,\chi)\|_{(X_{-,k})^*}^2 := \|(f,g)\|_{(Y_{-,k})^*}^2 + \alpha\|(\zeta,\chi)\|_{(Y_{-,k})^*}^2,$$

$$\|(f,g)\|_{(Y_{-,k})^*}^2 := \|f\|_{(U_{-,k})^*}^2 + \|g\|_{(P_{-,k})^*}^2,$$

$$\|f\|_{(U_{-,k})^*}^2 := h_k^2\|f\|_{H_0^1(\Omega)+\alpha^{-1/2}L^2(\Omega)}^2 \text{ and}$$

$$\|g\|_{(P_{-,k})^*}^2 := h_k^2\|g\|_{\alpha^{-1}L_0^2(\Omega)\cap\alpha^{-1/2}H_0^1(\Omega)}^2.$$

As stronger space, we choose $X_+ := Y_+ \times Y_+$, where $Y_+ := U_+ \times P_+$, $U_+ := [H^2(\Omega) \cap H_0^1(\Omega)]^d$ and $P_+ := H^1(\Omega) \cap L_0^2(\Omega)$, equipped with norms

$$\|x\|_{X_{+,k}}^2 := \|(u,p,\lambda,\mu)\|_{X_{+,k}}^2 := \|(u,p)\|_{Y_{+,k}}^2 + \alpha^{-1}\|(\lambda,\mu)\|_{Y_{+,k}}^2,$$

$$\|(u,p)\|_{Y_{+,k}}^2 := \|u\|_{U_{+,k}}^2 + \|p\|_{P_{+,k}}^2,$$

$$\|u\|_{U_{+,k}}^2 := h_k^2\|u\|_{H^1(\Omega)\cap\alpha^{1/2}H^2(\Omega)}^2 \text{ and}$$

$$\|p\|_{P_{+,k}}^2 := h_k^2\|p\|_{\alpha H^2(\Omega)+\alpha^{1/2}H^1(\Omega)}^2.$$

The additional conditions read as follows.

(A3) On all grid levels k, the approximation error result

$$\inf_{x_k \in X_k} \|x - x_k\|_X \lesssim \|x\|_{X_{+,k}} \qquad \text{for all } x \in X_+$$

is satisfied.

(A4) For all grid levels k, all $\mathcal{F} \in (X_-)^*$ the solution $x_\mathcal{F} \in X$ of the problem,

$$\text{find } x \in X \text{ such that} \qquad \mathcal{B}(x, \tilde{x}) = \mathcal{F}(\tilde{x}) \qquad \text{for all } \tilde{x} \in X, \qquad (10)$$

satisfies $x_\mathcal{F} \in X_+$ and the inequality

$$\|x_\mathcal{F}\|_{X_{+,k}} \lesssim \|\mathcal{F}\|_{(X_{-,k})^*}. \qquad (11)$$

Based on these assumptions, the following theorem shows the approximation property.

Theorem 1 *Let for $k = 0, 1, 2, \ldots$ the symmetric matrices $\mathcal{A}_k$ be obtained by discretizing problem (2) using a sequence of finite-dimensional nested subspaces $X_{k-1} \subseteq X_k \subset X$. Assume that there are Hilbert spaces $X_+ \subseteq X \subseteq X_-$ with mesh-dependent norms $\| \cdot \|_{X_{+,k}}$, $\| \cdot \|_X$ and $\| \cdot \|_{X_{-,k}}$ such that the conditions **(A1)**, **(A1a)**, **(A3)** and **(A4)** are satisfied. Then the coarse-grid correction (6) satisfies the approximation property*

$$\|x_k^{(1)} - x_k^*\|_{X_{-,k}} \leq C_A \sup_{\tilde{x}_k \in X_k} \frac{\mathcal{B}\left(x_k^{(0,\nu)} - x_k^*, \tilde{x}_k\right)}{\|\tilde{x}_k\|_{X_{-,k}}}, \tag{12}$$

where the constant C_A only depends on the constants that appear (implicitly) in the named conditions.

For a proof, see [16], Theorem 4.1.

Theorem 2 *Condition **(A3)** is satisfied.*

Proof This proof is analogous to the proof of Theorem 4.2 in [14]. However, to keep this paper as self-contained as possible, we give a proof of this theorem.

Note that it suffices to show approximation error results for the individual variables separately. Using a standard interpolation operator $\Pi_k : [H^2(\Omega)]^d \to U_k$, we obtain for the velocity field u

$$\|u - \Pi_k u\|_{L^2(\Omega)}^2 \lesssim h_k^2 \|u\|_{H^1(\Omega)}^2 \quad \text{and} \quad \|u - \Pi_k u\|_{H^1(\Omega)}^2 \lesssim h_k^2 \|u\|_{H^2(\Omega)}^2,$$

for all $u \in [H^2(\Omega)]^d$ and therefore

$$\inf_{u_k \in U_k} \|u - u_k\|_U^2 \leq \|u - \Pi_k u\|_U^2 = \|u - \Pi_k u\|_{L^2(\Omega)}^2 + \alpha^{1/2} \|u - \Pi_k u\|_{H^1(\Omega)}^2$$

$$\lesssim h_k^2 \left(\|u\|_{H^1(\Omega)}^2 + \alpha^{1/2} \|u\|_{H^2(\Omega)}^2 \right) = \|u\|_{U_{+,k}}^2.$$

The same can be done for the adjoined velocity λ. Also for the pressure distribution p we can do a similar estimate. The estimates

$$\inf_{p_k \in P_k} \|p - p_k\|_{L^2(\Omega)}^2 \lesssim h_k^2 \|p\|_{H^1(\Omega)}^2 \quad \text{and} \quad \inf_{p_k \in P_k} \|p - p_k\|_{H^1(\Omega)}^2 \lesssim h_k^2 \|p\|_{H^2(\Omega)}^2$$

are standard approximation error results which imply

$$
\inf_{p_k \in P_k} \|p - p_k\|_P^2 = \inf_{p_k \in P_k} \|p - p_k\|_{\alpha H^1(\Omega) + \alpha^{1/2} L^2(\Omega)}^2
$$

$$
= \inf_{\substack{p_k \in P_k \\ q_1 \in H^1(\Omega) \\ q_2 \in L^2(\Omega) \\ q_1 + q_2 = p - p_k}} \|q_1\|_{\alpha H^1(\Omega)}^2 + \|q_2\|_{\alpha^{1/2} L^2(\Omega)}^2
$$

$$
= \inf_{\substack{p_1 \in H^1(\Omega) \\ p_2 \in L^2(\Omega) \\ p_1 + p_2 = p}} \inf_{p_{1,k} \in P_k} \|p_1 - p_{1,k}\|_{\alpha H^1(\Omega)}^2 + \inf_{p_{2,k} \in P_k} \|p_2 - p_{2,k}\|_{\alpha^{1/2} L^2(\Omega)}^2
$$

$$
\leq \inf_{\substack{p_1 \in H^2(\Omega) \\ p_2 \in H^1(\Omega) \\ p_1 + p_2 = p}} \inf_{p_{1,k} \in P_k} \|p_1 - p_{1,k}\|_{\alpha H^1(\Omega)}^2 + \inf_{p_{2,k} \in P_k} \|p_2 - p_{2,k}\|_{\alpha^{1/2} L^2(\Omega)}^2
$$

$$
\lesssim h_k^2 \inf_{\substack{p_1 \in H^2(\Omega) \\ p_2 \in H^1(\Omega) \\ p_1 + p_2 = p}} \|p_1\|_{\alpha H^2(\Omega)}^2 + \|p_2\|_{\alpha^{1/2} H^1(\Omega)}^2 = h_k^2 \|p\|_{\alpha H^2(\Omega) + \alpha^{1/2} H^1(\Omega)}^2.
$$

The same can be done for the adjoined pressure μ. This finishes the proof. $\quad\square$

For showing **(A4)**, we recall Theorem 4.6 in [14] on the regularity of the generalized Stokes problem. For this purpose, we need a regularity assumption for the Poisson problem with homogeneous Neumann boundary conditions.

(R1) *Regularity of the Poisson problem.* Let $g \in L^2(\Omega)$ and $p \in H^1(\Omega) \cap L_0^2(\Omega)$ be such that

$$
(\nabla p, \nabla \tilde{p})_{H^1(\Omega)} = (g, \tilde{p})_{L^2(\Omega)} \text{ for all } \tilde{p} \in H^1(\Omega) \cap L_0^2(\Omega).
$$

Then $p \in H^2(\Omega)$ and $\|p\|_{H^2(\Omega)} \lesssim \|g\|_{L^2(\Omega)}$.

Such a regularity assumption can be guaranteed for convex polygonal domains (see, e.g., [6]).

Theorem 4.6 in [14] directly implies the following theorem.

Theorem 3 *Suppose that the regularity assumptions* **(R)** *and* **(R1)** *are satisfied. Let $f \in [L^2(\Omega)]^d$ and $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$. The solution of the problem, find $(u, p) \in Y$ such that*

$$
\alpha^{-1/2}(u, \tilde{u})_{L^2(\Omega)} + (\nabla u, \nabla \tilde{u})_{L^2(\Omega)} + (p, \nabla \cdot \tilde{u})_{L^2(\Omega)} = (f, \tilde{u})_{L^2(\Omega)}
$$
$$
(\nabla \cdot u, \tilde{p})_{L^2(\Omega)} = (g, \tilde{p})_{L^2(\Omega)}
$$

for all $(\tilde{u}, \tilde{p}) \in Y$, satisfies $(u, p) \in Y_+$ and the inequality

$$
\|u\|_{\alpha^{-1/2} H^1(\Omega) \cap H^2(\Omega)}^2 + \|p\|_{\alpha^{1/2} H^2(\Omega) + H^1(\Omega)}^2
$$
$$
\lesssim \|f\|_{\alpha^{1/2} H_0^1(\Omega) + L^2(\Omega)}^2 + \|g\|_{\alpha^{-1/2} L_0^2(\Omega) \cap H_0^1(\Omega)}^2
$$

is satisfied.

Proof We choose the parameter β (which occurs in [14]) to be $\beta := \alpha^{-1/2}$. $\quad\square$

Lemma 4 *Suppose that assumptions* **(R)** *and* **(R1)** *are satisfied. Let* $\mathcal{F} \in (X_-)^*$ *be arbitrarily but fixed. Then,* $x_{\mathcal{F}}$, *the solution of* (10), *satisfies* $x_{\mathcal{F}} \in X_+$ *and the bound*

$$\|x_{\mathcal{F}}\|_{X_{+,k}}^2 \lesssim \|\mathcal{F}\|_{(X_{-,k})^*}^2 + h_k^2 \left(\|u_{\mathcal{F}}\|_{H^1(\Omega)}^2 + \alpha^{-1} \|\lambda_{\mathcal{F}}\|_{H^1(\Omega)}^2 \right). \qquad (13)$$

Proof Let $\mathcal{F}(\tilde{u}, \tilde{p}, \tilde{\lambda}, \tilde{\mu}) := (f, \tilde{u})_{L^2(\Omega)} + (g, \tilde{p})_{L^2(\Omega)} + (\zeta, \tilde{\lambda})_{L^2(\Omega)} + (\chi, \tilde{\mu})_{L^2(\Omega)}$, where $f, \zeta \in [L^2(\Omega)]^d$ and $g, \chi \in H_0^1(\Omega) \cap L_0^2(\Omega)$.

Let $\hat{f} := f - u_{\mathcal{F}} + \alpha^{-1/2} \lambda_{\mathcal{F}}$ and $\hat{\zeta} := \zeta + \alpha^{-1} \lambda_{\mathcal{F}} + \alpha^{-1/2} u_{\mathcal{F}}$. Then we can rewrite the KKT-system as follows:

$$(\nabla \lambda_{\mathcal{F}}, \nabla \tilde{u})_{L^2(\Omega)} + \alpha^{-1/2} (\lambda_{\mathcal{F}}, \tilde{u})_{L^2(\Omega)} + (\mu_{\mathcal{F}}, \nabla \cdot \tilde{u})_{L^2(\Omega)} = (\hat{f}, \tilde{u})_{L^2(\Omega)}$$
$$(\nabla \cdot \lambda_{\mathcal{F}}, \tilde{p})_{L^2(\Omega)} = (g, \tilde{p})_{L^2(\Omega)}$$

and

$$(\nabla u_{\mathcal{F}}, \nabla \tilde{\lambda})_{L^2(\Omega)} + \alpha^{-1/2} (u_{\mathcal{F}}, \tilde{\lambda})_{L^2(\Omega)} + (p_{\mathcal{F}}, \nabla \cdot \tilde{\lambda})_{L^2(\Omega)} = (\hat{\zeta}, \tilde{\lambda})_{L^2(\Omega)}$$
$$(\nabla \cdot u_{\mathcal{F}}, \tilde{\mu})_{L^2(\Omega)} = (\chi, \tilde{p})_{L^2(\Omega)}.$$

As $\hat{f} \in [L^2(\Omega)]^d$, $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$, $\hat{\zeta} \in [L^2(\Omega)]^d$ and $\chi \in H_0^1(\Omega) \cap L_0^2(\Omega)$, we obtain using Theorem 3 that $x_{\mathcal{F}} \in X_+$ and the following bounds are satisfied:

$$\|\lambda_{\mathcal{F}}\|_{\alpha^{-1/2} H^1(\Omega) \cap H^2(\Omega)}^2 + \|\mu\|_{\alpha^{1/2} H^2(\Omega) + H^1(\Omega)}^2$$
$$\lesssim \|f - u_{\mathcal{F}} + \alpha^{-1/2} \lambda_{\mathcal{F}}\|_{\alpha^{1/2} H_0^1(\Omega) + L^2(\Omega)}^2 + \|g\|_{\alpha^{-1/2} L_0^2(\Omega) \cap H_0^1(\Omega)}^2$$

and

$$\|u_{\mathcal{F}}\|_{\alpha^{-1/2} H^1(\Omega) \cap H^2(\Omega)}^2 + \|p_{\mathcal{F}}\|_{\alpha^{1/2} H^2(\Omega) + H^1(\Omega)}^2$$
$$\lesssim \|\zeta - \alpha^{-1} \lambda_{\mathcal{F}} + \alpha^{-1/2} u_{\mathcal{F}}\|_{\alpha^{1/2} H_0^1(\Omega) + L^2(\Omega)}^2 + \|\chi\|_{\alpha^{-1/2} L_0^2(\Omega) \cap H_0^1(\Omega)}^2.$$

We can combine these two estimates and obtain

$$\|u_{\mathcal{F}}\|_{H^1(\Omega) \cap \alpha^{1/2} H^2(\Omega)}^2 + \|p_{\mathcal{F}}\|_{\alpha H^2(\Omega) + \alpha^{1/2} H^1(\Omega)}^2$$
$$+ \alpha^{-1} \|\lambda_{\mathcal{F}}\|_{H^1(\Omega) \cap \alpha 1/2 H^2(\Omega)}^2 + \alpha^{-1} \|\mu\|_{\alpha H^2(\Omega) + \alpha^{1/2} H^1(\Omega)}^2$$
$$\lesssim \|f\|_{H_0^1(\Omega) + \alpha^{-1/2} L^2(\Omega)}^2 + \|g\|_{\alpha^{-1} L_0^2(\Omega) \alpha^{-1/2} \cap H_0^1(\Omega)}^2$$
$$+ \alpha \|\zeta\|_{H_0^1(\Omega) + \alpha^{-1/2} L^2(\Omega)}^2 + \alpha \|\chi\|_{\alpha^{-1} L_0^2(\Omega) \cap \alpha^{-1/2} H_0^1(\Omega)}^2$$
$$+ \|u_{\mathcal{F}}\|_{H_0^1(\Omega) + \alpha^{-1/2} L^2(\Omega)}^2 + \alpha^{-1} \|\lambda_{\mathcal{F}}\|_{H_0^1(\Omega) + \alpha^{-1/2} L^2(\Omega)}^2.$$

Note that $\|u_{\mathcal{F}}\|_{H_0^1(\Omega) + \alpha^{-1/2} L^2(\Omega)} \leq \|u_{\mathcal{F}}\|_{H_0^1(\Omega)} = \|u_{\mathcal{F}}\|_{H^1(\Omega)}$ holds because of $u_{\mathcal{F}} \in [H_0^1(\Omega)]^d$. As the analogous holds also for $\lambda_{\mathcal{F}}$, this finishes the proof. $\quad\square$

To show condition **(A4)**, we have to bound $\|u_{\mathcal{F}}\|^2_{H^1(\Omega)} + \alpha^{-1}\|\lambda_{\mathcal{F}}\|^2_{H^1(\Omega)}$ from above. For showing such a result, we need some notation.

As $H^1_0(\Omega)$ is dense in $L^2(\Omega)$, for $u \in [H^2(\Omega)]^d$ the function $-\Delta u \in [L^2(\Omega)]^d$ can be approximated by some function $w^\epsilon \in [H^1_0(\Omega)]^d$ such that

$$\| - \Delta u - w^\epsilon \|^2_{L^2(\Omega)} \le \epsilon.$$

So, we can introduce an operator $-\Delta^\epsilon : [H^2(\Omega)]^d \to [H^1_0(\Omega)]^d$ such that

$$\| - \Delta u - (-\Delta^\epsilon)u \|^2_{L^2(\Omega)} \le \epsilon.$$

Analogously, we introduce the operator $\nabla^\epsilon : H^1(\Omega) \to [H^1_0(\Omega)]^d$ such that

$$\|\nabla p - \nabla^\epsilon p\|^2_{L^2(\Omega)} \le \epsilon.$$

Lemma 5 *Let $\mathcal{F} \in (X_-)^*$ and let $x_{\mathcal{F}} = (u_{\mathcal{F}}, p_{\mathcal{F}}, \lambda_{\mathcal{F}}, \mu_{\mathcal{F}})$ be the solution of (10). Then $x_{\mathcal{F}}$ satisfies the estimate*

$$h_k^2\left(\|u_{\mathcal{F}}\|^2_{H^1(\Omega)} + \alpha^{-1}\|\lambda_{\mathcal{F}}\|^2_{H^1(\Omega)}\right) \lesssim \|\mathcal{F}\|_{(X_{-,k})^*}\|x_{\mathcal{F}}\|_{X_{+,k}}. \qquad (14)$$

Proof Let $\mathcal{F}(\tilde{u}, \tilde{p}, \tilde{\lambda}, \tilde{\mu}) := (f, \tilde{u})_{L^2(\Omega)} + (g, \tilde{p})_{L^2(\Omega)} + (\zeta, \tilde{\lambda})_{L^2(\Omega)} + (\chi, \tilde{\mu})_{L^2(\Omega)}$, where $f, \zeta \in [L^2(\Omega)]^d$ and $g, \chi \in H^1_0(\Omega) \cap L^2_0(\Omega)$.

The idea of this proof is to show that for all $\epsilon > 0$ there is some $\tilde{x}^\epsilon \in X$ such that

$$\begin{aligned}
\mathcal{F}(\tilde{x}^\epsilon) &- \mathcal{B}(x_{\mathcal{F}}, \tilde{x}^\epsilon) \\
&\lesssim h_k^{-2}\|\mathcal{F}\|_{(X_{-,k})^*}\|x_{\mathcal{F}}\|_{X_{+,k}} - \|u_{\mathcal{F}}\|^2_{H^1(\Omega)} - \alpha^{-1}\|\lambda_{\mathcal{F}}\|^2_{H^1(\Omega)} \qquad (15) \\
&\quad + \epsilon(\alpha^{1/2} + \alpha^{-1/2})h_k^{-1}(\|x_{\mathcal{F}}\|_{X_{+,k}} + \|\mathcal{F}\|_{(X_{-,k})^*}) + \epsilon^2.
\end{aligned}$$

Note that the left-hand-side of the inequality is 0. Therefore, this would be sufficient to show the statement of the lemma, as $\epsilon > 0$ can be chosen arbitrarily small.

In the following, we show that (15) is satisfied for the choice $\tilde{x}^\epsilon := (-\Delta^\epsilon u_{\mathcal{F}}, -\nabla\cdot\nabla^\epsilon p_{\mathcal{F}}, \Delta^\epsilon \lambda_{\mathcal{F}}, \nabla\cdot\nabla^\epsilon \mu_{\mathcal{F}})$. We estimate the individual summands of $\mathcal{F}(\tilde{x}^\epsilon) - \mathcal{B}(x_{\mathcal{F}}, \tilde{x}^\epsilon)$ separately. For the first one, we obtain

$$\begin{aligned}
- (u_{\mathcal{F}}, -\Delta^\epsilon u_{\mathcal{F}})_{L^2(\Omega)} &\le -(u_{\mathcal{F}}, -\Delta u_{\mathcal{F}})_{L^2(\Omega)} + \epsilon\|u_{\mathcal{F}}\|_{L^2(\Omega)} \\
&= -(\nabla u_{\mathcal{F}}, \nabla u_{\mathcal{F}})_{L^2(\Omega)} + \epsilon\|u_{\mathcal{F}}\|_{L^2(\Omega)} \lesssim -\|u_{\mathcal{F}}\|^2_{H^1(\Omega)} + \epsilon h_k^{-1}\|x_{\mathcal{F}}\|_{X_{+,k}}
\end{aligned}$$

due to the fact that $u_{\mathcal{F}} \in [H^2(\Omega) \cap H^1_0(\Omega)]^d$ and due to Friedrichs' inequality. The same can be done for $\alpha^{-1}(\lambda_{\mathcal{F}}, \Delta^\epsilon \lambda_{\mathcal{F}})_{L^2(\Omega)}$.

For the next two summands,

$$\begin{aligned}
- (\nabla u_{\mathcal{F}}, \nabla \Delta^\epsilon \lambda_{\mathcal{F}})_{L^2(\Omega)} &- (\nabla \lambda_{\mathcal{F}}, \nabla(-\Delta^\epsilon) u_{\mathcal{F}})_{L^2(\Omega)} \\
&= (\Delta u_{\mathcal{F}}, \Delta^\epsilon \lambda_{\mathcal{F}})_{L^2(\Omega)} - (\Delta \lambda_{\mathcal{F}}, \Delta^\epsilon u_{\mathcal{F}})_{L^2(\Omega)} \\
&\le (\Delta u_{\mathcal{F}}, \Delta \lambda_{\mathcal{F}})_{L^2(\Omega)} - (\Delta \lambda_{\mathcal{F}}, \Delta u_{\mathcal{F}})_{L^2(\Omega)} + \epsilon(\|\Delta u_{\mathcal{F}}\|_{L^2(\Omega)} + \|\Delta \lambda_{\mathcal{F}}\|_{L^2(\Omega)}) \\
&\le \epsilon(\|u_{\mathcal{F}}\|_{H^2(\Omega)} + \|\lambda_{\mathcal{F}}\|_{H^2(\Omega)}) \le \epsilon h_k^{-1}(\alpha^{1/4} + \alpha^{-1/4})\|x_{\mathcal{F}}\|_{X_{+,k}}
\end{aligned}$$

is satisfied due to the fact that Δ^ϵ maps into $[H_0^1(\Omega)]^d$.

For the next two summands, we obtain

$$
\begin{aligned}
&-(\nabla \cdot u_\mathcal{F}, \nabla \cdot \nabla^\epsilon \mu_\mathcal{F})_{L^2(\Omega)} - (\nabla \cdot (-\Delta^\epsilon) u_\mathcal{F}, \mu_\mathcal{F})_{L^2(\Omega)} \\
&= (\nabla \nabla \cdot u_\mathcal{F}, \nabla^\epsilon \mu_\mathcal{F})_{L^2(\Omega)} - (\Delta^\epsilon u_\mathcal{F}, \nabla \mu_\mathcal{F})_{L^2(\Omega)} \\
&\leq (\nabla \nabla \cdot u_\mathcal{F}, \nabla^\epsilon \mu_\mathcal{F})_{L^2(\Omega)} - (\Delta^\epsilon u_\mathcal{F}, \nabla^\epsilon \mu_\mathcal{F})_{L^2(\Omega)} + \epsilon \|\Delta^\epsilon u_\mathcal{F}\|_{L^2(\Omega)} \\
&\leq -(\nabla u_\mathcal{F}, \nabla \nabla^\epsilon \mu_\mathcal{F})_{L^2} - (\Delta u_\mathcal{F}, \nabla^\epsilon \mu_\mathcal{F})_{L^2} + \epsilon(\|\Delta u_\mathcal{F}\|_{L^2} + \|\nabla^\epsilon p_\mathcal{F}\|_{L^2} + \epsilon) \\
&= -(\nabla u_\mathcal{F}, \nabla \nabla^\epsilon \mu_\mathcal{F})_{L^2} + (\nabla u_\mathcal{F}, \nabla \nabla^\epsilon \mu_\mathcal{F})_{L^2} + \epsilon(\|\Delta u_\mathcal{F}\|_{L^2} + \|\nabla^\epsilon p_\mathcal{F}\|_{L^2} + \epsilon) \\
&\leq \epsilon(\|u_\mathcal{F}\|_{H^2(\Omega)} + \|p_\mathcal{F}\|_{H^1(\Omega)} + 2\epsilon) \lesssim \epsilon h_k^{-1}(\alpha^{-1/4} + \alpha^{-1/2})\|x_\mathcal{F}\|_{X_{+,k}} + \epsilon^2.
\end{aligned}
$$

The same can be done for $-(\nabla \cdot \lambda_\mathcal{F}, -\nabla \cdot \nabla^\epsilon p_\mathcal{F})_{L^2(\Omega)} - (\nabla \cdot \Delta^\epsilon \lambda_\mathcal{F}, p_\mathcal{F})_{L^2(\Omega)}$.

Let $f_2 \in [H_0^1(\Omega)]^d$ and $f_1 := f - f_2$. Then

$$
(f_1, -\Delta^\epsilon u_\mathcal{F})_{L^2(\Omega)} \lesssim \|f_1\|_{\alpha^{-1/2}L^2(\Omega)} \|u_\mathcal{F}\|_{\alpha^{1/2}H^2(\Omega)} + \epsilon \alpha^{1/4} \|f_1\|_{\alpha^{-1/2}L^2(\Omega)}
$$

holds as well as

$$
\begin{aligned}
(f_2, -\Delta^\epsilon u_\mathcal{F})_{L^2(\Omega)} &\lesssim (\nabla f_2, \nabla u_\mathcal{F})_{L^2(\Omega)} + \epsilon \|f_2\|_{L^2(\Omega)} \\
&\lesssim \|f_2\|_{H^1(\Omega)} \|u_\mathcal{F}\|_{H^1(\Omega)} + \epsilon \|f_2\|_{H^1(\Omega)}.
\end{aligned}
$$

This implies

$$
\begin{aligned}
(f, -&\Delta^\epsilon u)_{L^2(\Omega)} \\
&\lesssim \|f\|_{H_0^1(\Omega) + \alpha^{-1/2}L^2(\Omega)} \|u\|_{H^1(\Omega) \cap \alpha^{1/2}H^2(\Omega)} \\
&\quad + \epsilon(1 + \alpha^{1/4})\|f\|_{H_0^1(\Omega) + \alpha^{-1/2}L^2(\Omega)} \\
&\lesssim \|f\|_{H_0^1(\Omega) + \alpha^{-1/2}L^2(\Omega)} \|u\|_{H^1(\Omega) \cap \alpha^{1/2}H^2(\Omega)} + \epsilon h_k^{-1}(1 + \alpha^{1/4})\|\mathcal{F}\|_{(X_{-,k})^*}.
\end{aligned}
$$

Let $p_2 \in H^2(\Omega)$ and $p_1 := p_\mathcal{F} - p_2 \in H^1(\Omega)$. We have

$$
\begin{aligned}
(g, -\nabla \cdot \nabla^\epsilon p_1)_{L^2(\Omega)} &= (\nabla g, \nabla^\epsilon p_1)_{L^2(\Omega)} \\
&\lesssim \|g\|_{\alpha^{-1/2}H^1(\Omega)} \|p_1\|_{\alpha^{1/2}H^1(\Omega)} + \epsilon \|g\|_{H^1(\Omega)}.
\end{aligned}
$$

Moreover, using $g \in H_0^1(\Omega)$, we have also

$$
\begin{aligned}
(g, -\nabla \cdot \nabla^\epsilon p_2)_{L^2(\Omega)} &= (\nabla g, \nabla^\epsilon p_2)_{L^2(\Omega)} \lesssim (\nabla g, \nabla p_2)_{L^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)} \\
&= -(g, \nabla \cdot \nabla p_2)_{L^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)} \leq \|g\|_{\alpha^{-1}L^2(\Omega)} \|p_2\|_{\alpha H^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)}
\end{aligned}
$$

and therefore

$$
\begin{aligned}
(g, -&\nabla \cdot \nabla^\epsilon p_\mathcal{F})_{L^2(\Omega)} \\
&\lesssim \|g\|_{\alpha^{-1}L^2(\Omega) \cap \alpha^{-1/2}H^1(\Omega)} \|p_\mathcal{F}\|_{\alpha^{1/2}H^1(\Omega) + \alpha H^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)} \\
&\lesssim \|g\|_{\alpha^{-1}L^2(\Omega) \cap \alpha^{-1/2}H^1(\Omega)} \|p_\mathcal{F}\|_{\alpha^{1/2}H^1(\Omega) + \alpha H^2(\Omega)} + \epsilon h_k^{-1}\alpha^{1/4}\|\mathcal{F}\|_{(X_{-,k})^*}
\end{aligned}
$$

is satisfied. The same can be done for $(\zeta, \Delta^\epsilon \lambda_\mathcal{F})_{L^2(\Omega)}$ and $(\chi, \nabla \cdot \nabla^\epsilon \mu_\mathcal{F})_{L^2(\Omega)}$.

Combining these results, we immediately obtain (15), which finishes the proof. $\qquad\square$

Theorem 4 *Condition* **(A4)** *is satisfied.*

Proof By combining (13) and (14), we obtain

$$\|x_{\mathcal{F}}\|_{X_{+,k}}^2 \le C \left(\|\mathcal{F}\|_{(X_{-,k})^*}^2 + \|\mathcal{F}\|_{(X_{-,k})^*} \|x_{\mathcal{F}}\|_{X_{+,k}} \right)$$

for some constant $C > 0$ (independent of k and β) which implies

$$\|x_{\mathcal{F}}\|_{X_{+,k}} \le \frac{1}{2} \left(C + \sqrt{4C + C^2} \right) \|\mathcal{F}\|_{(X_{-,k})^*},$$

i.e. (11), which finishes the proof. $\square$

So, we have shown condition **(A4)**. So, Theorem 1 implies the approximation property. Note, that we have now shown the approximation property in the norm $\|\cdot\|_{X_{-,k}}$, i.e., (12). The next step is to show the approximation property in the norm-pair $\|\!|\cdot|\!\|_{0,k}$ and $\|\!|\cdot|\!\|_{2,k}$, i.e., (9).

To show (9), the following lemma is sufficient.

Lemma 6 *The inequality*

$$\|\!|x_k|\!\|_{0,k} \lesssim \|x_k\|_{X_{-,k}} \tag{16}$$

is satisfied for all $x_k \in X_k$.

Proof The proof of this lemma is based on Lemma 4.7 in [14]. Lemma 4.7 states (in the notation of the present paper and for the choice $\beta := \alpha^{-1/2}$) that

$$\alpha^{-1/2} \|\!|y_k|\!\|_{Y,0,k}^2 \lesssim \alpha^{-1/2} \|y_k\|_{Y_{-,k}}^2$$

is satisfied for all $y_k \in Y_k$. As for $x_k = (y_k, \psi_k) \in X_k = Y_k \times Y_k$ both,

$$\|\!|(y_k, \psi_k)|\!\|_{0,k}^2 = \|\!|y_k|\!\|_{Y,0,k}^2 + \alpha^{-1} \|\!|\psi_k|\!\|_{Y,0,k}^2$$

and

$$\|(y_k, \psi_k)\|_{X_{-,k}}^2 = \|y_k\|_{Y_{-,k}}^2 + \alpha^{-1} \|\psi_k\|_{Y_{-,k}}^2,$$

is satisfied by definition, (16) follows immediately. $\square$

So, we have shown the approximation property (9). So, we obtain the following overall convergence result.

Theorem 5 *Assume that*

– *the regularity assumptions* **(R)** *and* **(R1)** *are satisfied on the domain Ω,*
– *the problem is discretized using the Taylor-Hood element and*
– *the normal equation smoother introduced above is used as smoother.*

Then the two-grid method converges if sufficiently many smoothing steps are applied, i.e., we have

$$\|\!|x_k^{(1)} - x_k^*|\!\|_{0,k} \le q(\nu) \|\!|x_k^{(0)} - x_k^*|\!\|_{0,k},$$

with $q(\nu) := C_S\, C_A\, \nu^{-1/2}$, where the constants C_A and C_S are independent of the grid level k and the choice of the parameter α.

Note that we have shown that the method converges in the norm $\|\|\cdot\|\|_{0,k}$ if sufficiently many pre-smoothing steps are applied. The application of post-smoothing steps does not derogate the convergence because the proposed smoother is power-bounded. Moreover, if we assume that only post-smoothing steps are applied, the combination of smoothing property and approximation property (which now have to be combined in the inverse order) leads to convergence in the residual norm $\|\|\cdot\|\|_{2,k}$. Again, due to power-boundedness of the smoother, the method stays convergent if, besides sufficiently many post-smoothing steps, also pre-smoothing steps are applied.

For all the mentioned cases, the convergence of the W-cycle multigrid method follows under weak assumptions, cf. [7].

5 Numerical Results

In this section, we illustrate the convergence theory presented within this paper with numerical experiments.

The domain Ω was chosen to be the unit square $\Omega := (0,1)^2$. As mentioned in Section 2, the weak inf-sup-condition **(S)** can be shown for the Taylor-Hood element only if at least one vertex of each element is located in the interior of the domain Ω. As this is not satisfied for the standard decomposition of the unit square into two triangular elements, we have chosen the coarsest grid (grid level $k = 0$) to be a decomposition of the domain Ω into 8 triangles, cf. Fig. 1. The grid levels $k = 1, 2, \ldots$ were constructed by uniform refinement, i.e., every triangle was decomposed into four subtriangles.

The desired velocity field (desired state) u_D was chosen to be

$$u_D(\xi_1, \xi_2) := \begin{cases} \begin{pmatrix} \xi_2 - \frac{1}{2} \\ \frac{1}{2} - \xi_1 \end{pmatrix} & \text{for } \sqrt{\left(\xi_1 - \frac{1}{2}\right)^2 + \left(\xi_2 - \frac{1}{2}\right)^2} < \frac{4}{5} \\ 0 & \text{otherwise.} \end{cases}$$

The desired velocity field is visualized in the left-hand-side picture in both, Fig. 2 and 3.

For solving the discretized KKT-system, we have used the proposed W-cycle multigrid method. We have applied ν pre- and ν post-smoothing steps using the normal equation smoother. The matrix $\mathcal{L}_k$ was chosen as follows

$$\mathcal{L}_k := \begin{pmatrix} \hat{A}_k & & & \\ & \hat{S}_k & & \\ & & \alpha^{-1}\hat{A}_k & \\ & & & \alpha^{-1}\hat{S}_k \end{pmatrix}, \tag{17}$$

where $\hat{A}_k := \mathrm{diag}\left(M_{U,k} + \alpha^{1/2}K_{U,k}\right)$ and $\hat{S}_k := \alpha\,\mathrm{diag}\left(D_k\hat{A}_k^{-1}D_k^T\right)$. Here, $M_{U,k}$ and $K_{U,k}$ are the mass matrix and the stiffness matrix, representing the L^2-inner product and the H^1-inner product in U_k, respectively. The matrix D_k represents the bilinear form $d(u_k, p_k) = (\nabla \cdot u_k, p_k)_{L^2(\Omega)}$ on $U_k \times P_k$. Note that the matrix $\mathcal{L}_k$, introduced above, is spectrally equivalent to the matrix $\mathcal{L}_k$,

introduced in Section 3. Therefore, the choice proposed above is also covered
by the convergence theory. The damping parameter was chosen to be $\tau = 0.35$
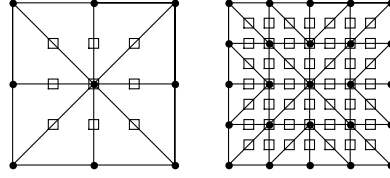for all grid levels k and all choices of α.

Fig. 1 Discretization on grid levels $k = 1$ and $k = 2$, where the squares denote the degrees
of freedom of (the components of) u and λ are the the dots denote the degrees of freedom
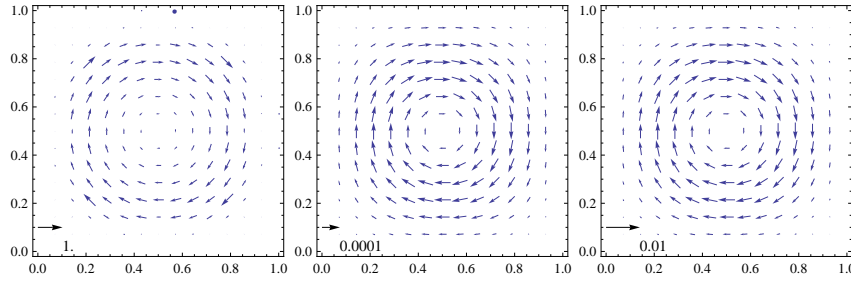of p and μ

Fig. 2 Desired velocity field u_D, optimal velocity field u and optimal control f for $\alpha = 1$
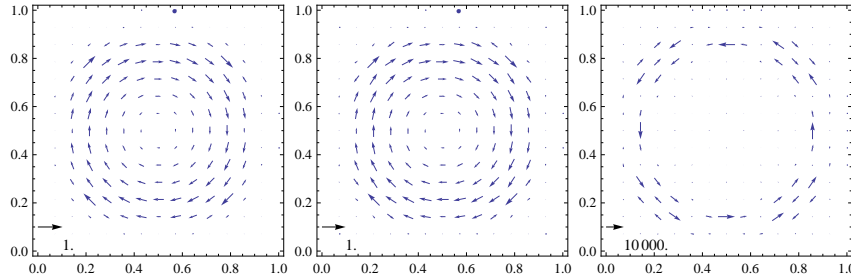on grid level $k = 3$

Fig. 3 Desired velocity field u_D, optimal velocity field u and optimal control f for $\alpha = 10^{-12}$ on grid level $k = 3$

The solution of the optimal control problem can be seen in Fig. 2 and
3. Note that the desired velocity field is a general L^2-function (due to the

jump) and therefore it cannot be reached by the (optimal) velocity field which is an H^1-function and therefore continuous. For the case $\alpha = 1$, we observe the optimal velocity field and the control to be rather smooth. (The control does not take large values). For small values of α, like $\alpha = 10^{-12}$, the desired velocity field is approximated quite well, cf. Fig. 3. This is achieved by rather large values of the control (high forces) which are concentrated on the region where the desired velocity field has its jump. Forces have to be applied with the same orientation as the desired state, as well as with the opposite orientation. (In the picture mainly the forces with opposite orientation can be seen.) As mentioned in the introduction, we are interested in a fast linear solver which also works well for such small choices of α.

The number of iterations and the convergence rate were measured as follows: we start with $x_k^{(0)} = 0$ and measure the reduction of the error in each step using the residual norm $\| \cdot \|_{2,k}$. The iteration was stopped when the initial error was reduced by a factor of $\epsilon = 10^{-6}$. The convergence rates q is the mean convergence rate in this iteration, i.e.,

$$q = \left(\frac{\|x_k^{(n)} - x_k^*\|_{2,k}}{\|x_k^{(0)} - x_k^*\|_{2,k}} \right)^{1/n},$$

where n is the number of iterations needed to reach the stopping criterion. Here, x_k^* is the exact solution and $x_k^{(i)}$ is the i-th iterate.

$\nu = 1+1$		$\nu = 2+2$		$\nu = 4+4$		$\nu = 8+8$	
n	q	n	q	n	q	n	q
61	0.796	32	0.647	21	0.507	15	0.390

Table 1 Number of iterations n and convergence rate q depending on $\nu = \nu_{pre} + \nu_{post}$, the number of pre- and post-smoothing steps, on grid level $k = 4$ for $\alpha = 1$

	$\alpha = 1$		$\alpha = 10^{-3}$		$\alpha = 10^{-6}$		$\alpha = 10^{-9}$		$\alpha = 10^{-12}$	
	n	q	n	q	n	q	n	q	n	q
$k = 3$	32	0.648	33	0.651	35	0.673	48	0.749	51	0.760
$k = 4$	32	0.647	32	0.646	33	0.657	46	0.738	73	0.827
$k = 5$	32	0.645	32	0.644	32	0.644	39	0.697	60	0.793
$k = 6$	31	0.636	31	0.636	31	0.635	32	0.647	46	0.739
$k = 7$	29	0.620	29	0.620	29	0.618	29	0.621	42	0.716

Table 2 Number of iterations n and convergence rate q for $\nu = 2 + 2$ pre- and post-smoothing steps

In Table 1 we compare for a fixed grid level (level $k = 4$) and a fixed choice $\alpha = 1$ the convergence rates for several choices of ν, the number of pre- and post-smoothing steps. We see that the convergence rates behave approximately like $\nu^{-1/2}$. This is consistent with the theory which guarantees the convergence rate being bounded by $C\,\nu^{-1/2}$ as this only describes the asymptotic behavior.

In Table 2 we compare various grid levels k and choices of the parameter α. Here, we have used a fixed choice of $\nu = 2 + 2$ pre- and post-smoothing steps. First we observe that the number of iterations seems to be well-bounded for all grid levels k which yields an optimal convergence behavior. Moreover, we see that the number of iterations is also well-bounded for a wide range of choices of the parameter α, i.e., we observe also robust convergence as predicted by the convergence theory.

It has to be mentioned that for the model problem, also the (more efficient) V-cycle multigrid method converges with rates comparable to the convergence rates of the W-cycle multigrid method. However, the V-cycle is not covered by the convergence theory.

6 Conclusions and Further Work

In the present paper we have shown that the construction of an all-at-once multigrid method for a Stokes control problem is possible. Here, a preconditioned normal equation smoother was chosen. The overall numerical complexity of this method seems to be comparable to block-preconditioned MINRES iterations, cf., e.g., Table 4.2 in [18], which shows the number of MINRES iterations needed. (Note that in each MINRES step one multigrid cycle is applied to each component of the overall block-matrix, i.e., to the velocity u, the pressure p, the adjoined velocity λ and the adjoined pressure μ.)

One advantage of the all-at-once multigrid method, introduced in the present paper, is the fact that an outer iteration is not necessary, the multigrid iteration is an linear iteration scheme which can be directly applied to solve the problem. As we could show the approximation property for a particular choice of norms, the construction of other smoothers is of particular interest. The convergence rates we have observed in this paper for a multigrid method with normal equation smoothing are comparable with the convergence rates observed in [16] for a multigrid method with normal equation smoothing applied to an optimal control problem with elliptic state equation. For that problem we have seen that other smoothers are available which lead to much faster convergence rates, cf. [15] and others. Similar improvements were possible for the generalized Stokes problem, cf. [14]. Therefore, it seems to be reasonable to construct faster smoothers also for the the Stokes control problem.

References

1. R. Adams and J. Fournier. *Sobolev Spaces*. Academic Press, 2008. 2nd ed.
2. M. Bercovier and O. Pironneau. Error estimates for finite element method solution of the Stokes problem in primitive variables. *Numerische Mathematik*, 33:211 − 224, 1979.
3. J. Bergh and J. Löfström. *Interpolation Spaces, an Introduction*. Springer, Berlin, 1976.
4. A. Borzi and V. Schulz. Multigrid Methods for PDE Optimization. *SIAM Review*, 51:361 − 395, 2009.
5. S.C. Brenner. Multigrid methods for parameter dependent problems. *RAIRO, Modélisation Math. Anal. Numér*, 30:265 − 297, 1996.
6. M. Dauge. Elliptic boundary value problems on corner domains. Smoothness and asymptotics of solutions. *Lecture Notes in Mathematics, 1341. Berlin etc.: Springer-Verlag*, 1988.
7. W. Hackbusch. *Multi-Grid Methods and Applications*. Springer, Berlin, 1985.
8. R.B. Kellogg and J.E Osborn. A regularity result for the Stokes problem in a convex polygon. *Journal of Functional Analysis*, 21(4):397–431, 1976.
9. M. Kollmann and W. Zulehner. A Robust Preconditioner for Distributed Optimal Control for Stokes Flow with Control Constraints. In Andrea Cangiani, Ruslan L. Davidchack, Emmanuil Georgoulis, Alexander N. Gorban, Jeremy Levesley, and Michael V. Tretyakov, editors, *Numerical Mathematics and Advanced Applications 2011*, pages 771–779. Springer Berlin Heidelberg, 2013.
10. M. Olshanskii. Multigrid Analysis for the Time Dependent Stokes Problem. *Mathematics of Computation*, 81(277):57 − 79, 2012.
11. M. Olshanskii, J. Peters, and A. Reusken. Uniform preconditioners for a parameter dependent saddle point problem with application to generalized stokes interface equations. *Numerische Mathematik*, pages 159–191, 2005.
12. J. W. Pearson. On the Role of Commutator Arguments in the Development of Regularization-Robust Preconditioners for Stokes Control Problems, 2012. submitted.
13. J. Schöberl, R. Simon, and W. Zulehner. A Robust Multigrid Method for Elliptic Optimal Control Problems. *SIAM J. on Numerical Analysis*, 49:1482 − 1503, 2011.
14. S. Takacs. A multigrid method for the time-dependent Stokes problem, 2013. Submitted.
15. S. Takacs and W. Zulehner. Convergence Analysis of Multigrid Methods with Collective Point Smoothers for Optimal Control Problems. *Computing and Visualization in Science*, 14(3):131–141, 2011.
16. S. Takacs and W. Zulehner. Convergence analysis of all-at-once multigrid methods for elliptic control problems under partial elliptic regularity. *SIAM J. on Numerical Analysis*, 2012. Accepted.
17. R. Verfürth. Error estimates for a mixed finite element approximation of the Stokes equations. *RAIRO*, 18:175 − 182, 1984.
18. W. Zulehner. Non-standard Norms and Robust Estimates for Saddle Point Problems. *SIAM J. on Matrix Anal. & Appl*, 32:536 − 560, 2011.